DOCUMENT RESUME

ED 163 066 TM 008 128

AUTHOR	Abrami, Philip C.; And Others
TITLE	Teacher Grading Standards and Student Ratings.
PUB DATE	Sep 78
NOTE	17p.; Paper presented at the Annual Meeting of the American Psychological Association (81st, Toronto, Ontario, Canada, September, 1978)
AVAILABLE FROM	Philip C. Abrami, Centre for Learning and Development, McGill University, Montreal, Quebec, Canada H3A 2K6
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Academic Standards; College Teachers; Evaluation Criteria; Foreign Countries; *Grades (Scholastic); Grading; Higher Education; *Predictor Variables; Student Attitudes; *Student Evaluation of Teacher Performance; *Student Motivation; *Teaching Styles
IDENTIFIERS	Canada

ABSTRACT

Two experiments were conducted to determine the effects of teacher grading standards on student ratings of teacher performance. In experiment one, three teacher characteristics were factorially manipulated: instructor expressiveness (high, low), lecture content (high, low), and grading standards (B, C+, C). In experiment two, student incentive (high, low), instructor expressiveness (high, low), and grading standards (B+, B, C+, C) were manipulated. A complex relationship between grading standards and student ratings were found. Under certain conditions, assigning high grades improved ratings; while under other conditions, high grades either adversely affected ratings or did not affect them at all. (Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163066 -

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Teacher Grading Standards and Student Ratings

Philip C. Abrami, Wenda J. Dickens, Raymond P. Perry,

and Les Leventhal

University of Manitoba

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Philip Abrami

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM008 128

Running Head: Teacher Grading Standards

Paper presented at the annual meeting of the American Psychological Association, Toronto, September, 1978.

Requests for reprints should be addressed to: Philip C. Abrami, Centre for Learning and Development, McGill University, Montreal, Quebec, Canada H3A 2K6.

Abstract

Two experiments were conducted to determine the effects of teacher grading standards on student ratings. In experiment one, we factorially manipulated three teacher characteristics: instructor expressiveness (high, low), lecture content (high, low), and grading standards (B, C+, C). In experiment two, we manipulated student incentive (high, low), instructor expressiveness (high, low), and grading standards (B+, B, C+, C). We found a complex relationship between grading standards and ratings. Under certain conditions, assigning high grades improved ratings while under other conditions high grades either adversely affected ratings or did not affect them at all.

Teacher Grading Standards and Student Ratings

Many teacher rating form (TRF) researchers are interested in how student opinions about teachers are formed. These researchers may criticize TRFs if grades students receive affect ratings. However, this criticism hold only when certain factors which are theoretically unrelated to teacher effectiveness produce the grades/ratings relationship. There are at least three ways that a grades/ratings relationship may occur. For example, differences in teacher standards for assigning grades can produce class-to-class differences in grades and classes receiving high grades may award high ratings in gratitude. Here, finding that grades affect ratings would show that instructor grading standards--thought to be unrelated to teacher effectiveness--affect ratings. On the other hand, class-to-class differences in grades may reflect differences in instructor effectiveness and classes with higher grades may award better ratings. Here, finding that grades affect ratings would show that ratings identify effective teachers. Finally, within a class, student-to-student differences in grades may reflect differences in a student characteristic (e.g., ability) and students with better grades may award higher ratings. Here, finding that grades affect ratings would show that a student characteristic affects ratings. Saying that grades affect ratings could mean that: (a) grading standards affect ratings, or (b) effective instructors affect ratings, or (c) student characteristics affect ratings, or (d) some combination of factors affect ratings. So the real question for us was not simply whether grades affect ratings but something more specific. We decided to investigate whether instructors receive different ratings based on the grading standards they employ.

Evidence on the grades/ratings relationship is available from many studies. Feldman's (1976) extensive review of the grades/ratings literature describes separately the studies which employed: (a) students within a class as the units of analysis, (b) students pooled across two or more classes as the units of analysis, and (c) classes as the units of analysis. These different units of analysis when applied to the same data do not allow for similar interpretations (see Leventhal, Abrami, & Perry, 1977). For example, when students within a single class are the units of analysis, variability in ratings may be explained by student-to-student differences. However, when students pooled across two or more classes are the units of analysis, variability in ratings may be explained by student, teacher, and situational differences. Choosing the appropriate unit of analysis depends on the situation of interest. In the present study, we were interested in the effects of a teacher characteristic--grading standards--on ratings, so the class, with students randomly assigned, seemed to be the appropriate unit of analysis.

Correlational evidence is available from many studies, employing the class as the unit of analysis, investigating the relationship between TRF ratings and grades (see Feldman, 1976). Unfortunately, Leventhal (1975) has pointed out that, since few of these studies randomly assigned students to classes, relationships found between ratings and grades may be explained by many rival causal hypotheses. For example, a positive relationship will result if motivated students seek out certain teachers, work harder, and rate teachers higher than unmotivated students.

Only Snyder and Clair's (1976) study allows for this relationship to be explained exclusively by teacher differences in grading standards. Snyder and

Clair (1976) used in the class, with students randomly assigned, as the unit of analysis and manipulated grades while controlling other teacher differnces. They found that grades did affect teacher ratings: higher grades produced more positive ratings. Unfortunately, this study has limited generalizability. Students were tested on a single 10-minute tape recorded lecture and grades were randomly assigned to students. This differs from actual classes where grades within a class usually depend on student performances. Therefore, the purpose of the first experiment was to determine if grading standards affect ratings for different teachers when grades within a class depend on student performance.

EXPERIMENT 1

Method

Subjects. The subjects were 136 male and female students from the 1976-77 Introductory Psychology course at the University of Manitoba who received course credit for research participation.

Materials. The lectures were the "Doctor Fox" videotapes previously employed by Williams and Ware (1977). Four black and white videotaped lectures were prepared on each of two topics, the biochemistry of learning and effects of environmental stimulation on the brain. The tapes for a topic varied on two dimensions: instructor expressiveness (high, low) and lecture content (high, low). Instructor expressiveness was manipulated by varying the lecturer's enthusiasm, humor, physical movement, etc. Lecture content was manipulated by varying the number of teaching points contained in the lecture.

The students evaluated the instructor on the 18-item Pohlmann (1975) TRF. Williams and Ware's multiple choice quizzes were used with the

biochemistry of learning videotapes (26-items) and the effects of environmental stimulation on the brain videotapes (20-items).

Experimental design. Instructor expressiveness (low, high), lecture content (low, high), and grading standards (B, C+, C), were combined to form a 2 x 2 x 3 between groups factorial design. All students saw two lectures, two weeks apart. The dependent variables were an averaged TRF rating (Pohlmann, 1975) and a total correct quiz score.

Procedure. The first lecture was one of the four biochemistry of learning videotapes, presented on an Electrohome 21-inch (53 cm) television monitor. Following the first lecture, students completed a TRF and quiz. After two weeks, students returned for the second lecture. At this time the grading standards manipulation was introduced. Students were told that "in order to replicate typical classroom conditions where students receive their test results, the quizzes [from the first lecture] have been scored and the results are being distributed to you." Feedback to students on quiz results consisted of: (a) their total number correct, (b) the class average correct, (c) their letter grade, and (d) the average class letter grade. Three standards for awarding letter grades were employed--B, C+, C. Each of these grades represented the average letter grade for students in an experimental session viewing a particular lecture. Individual student grades were assigned on the basis of the deviation of individual quiz results from the average for that experimental session. A student whose quiz score was one standard deviation above the average received one letter grade higher than the average.

Following the distribution of the quiz results students viewed the second lecture which was one of the four environmental stimulation on the

brain videotapes. Students were shown the videotape which represented the same level of Expressiveness and Content as the first lecture. Following the lecture students again completed the TRF and appropriate quiz.

Results and Discussion

To determine whether grading standards affected ratings we analyzed the TRF scores for the second lecture in an Expressiveness x Content x Standards analysis of variance (ANOVA) using computer program BMDP2V (Dixon, 1975). The results of the ANOVA showed no main effect of Standards, $F(2, 124) = 1.72$, $p > .05$. There was, however, a significant Expressiveness x Standards interaction, $F(2, 124) = 5.51$, $p < .05$, which accounted for 5.0% of the variance in ratings (determined by computing omega squared [Hays, 1973, pp. 512-514]). This interaction is illustrated in Figure 1. The simple main effects of

- - - - - - - - - - - - - - -
Insert Figure 1 about here
- - - - - - - - - - - - - - -

Standards were significant, both at low expressiveness, $F(2, 124) = 3.29$, $p < .05$, and at high expressiveness, $F(2, 124) = 3.78$, $p < .05$. For the high expressiveness conditions, pairwise comparisons revealed that the lowest grade group (C) gave worse ratings than either the C+ group, $t(124) = 2.77$, $p < .05$, or the B group, $t(124) = 2.09$, $p < .05$. For the low expressiveness conditions, the highest grade group (B) gave worse ratings than either the C+ group, $t(124) = 2.19$, $p < .05$, or the C group, $t(124) = 2.16$, $p < .05$. In addition, neither the Content x Standards interaction, $F(2, 124) < 1.00$, $p > .05$, nor the Content x Expressiveness x Standards interaction, $F(2, 124) = 1.83$, $p > .05$, was significant.

The significant Expressiveness x Standards interaction found in Experiment 1 suggests that grading standards affected ratings. The results for the high

expressiveness conditions appear consistent with the findings of Snyder and Clair (1976): the lowest grade group gave the worst ratings. However, lenient grades did not insure the lecturer a good rating; when instructor expressiveness was low, the group that received the highest grades gave the worst ratings!

Surprised by these findings, we decided to conduct a second experiment First, we wanted to replicate these results with a different set of videotaped lectures. Second, we wanted to know how grades would affect ratings when student incentive to achieve is high. High student incentive probably distinguishes simulated classes from many typical classes and may alter the effect of grading students on ratings. For example, when incentive is high, the grades students receive may determine teacher ratings, regardless of instructor expressiveness. Finally, we added a fourth level of Standards (B+) to study the effect of very liberal grading on ratings.

EXPERIMENT 2

Method

Subjects. The subjects were 278 male and female students from the 1977-78 Introductory Psychology course at the University of Manitoba.

Materials. The lectures were color videotapes on each of two topics: repression and sex roles. For each topic, only instructor expressiveness (high, low) was manipulated since Content did not interact with Standards in Experiment 1.

Once again, the students evaluated the instructor on the 18-item Pohlmann (1975) TRF. Two 30-item multiple choice quizzes were constructed for use with the repression and sex roles videotapes.

Experimental design. Instructor expressiveness (low, high), student

incentive (low, high), and grading standards (B+, B, C+, C) were combined to form a 2 x 2 x 4 between groups factorial design. All students saw two lectures, two weeks apart. The dependent variables were an averaged TRF rating (Pohlmann, 1975) and a total correct quiz score.

Procedure. In this experiment, we used a 7-foot (2.2 meters) television projector (Advent 1000A) rather than a small monitor to reduce the artificiality of the lectures. For the low incentive students, running procedure was similar to Experiment 1. High incentive students were instructed prior to the first lecture (repression) that an average performance of B or better on the first and second quiz would result in their receiving two additional experimental credits.

Results and Discussion

To determine whether grading standards affected ratings we analyzed only the second lecture TRF scores in an Expressiveness x Incentive x Standards ANOVA. Means for these data are illustrated in Figure 2. We computed the

- - - - - - - - - - - - - - - -
Insert Figure 2 about here
- - - - - - - - - - - - - - - -

Expressiveness x Standards interaction at low incentive (i.e., a simple interaction test; Kirk, 1968, pp. 222-224) as a comparison to Experiment 1. This interaction was not significant, $F(3, 262) = 2.43$, $p > .05$. As a further comparison, we tested the Standards effect at both levels of Expressiveness. For low incentive students, grading standards affected ratings at high expressiveness, $F(3, 262) = 3.73$, $p < .05$, but not at low expressiveness, $F(3, 262) < 1.00$, $p > .05$. Pairwise comparisons among high expressiveness means revealed that B was significantly different from both C+, $t(262) = 3.28$, $p < .05$, and C, $t(262) = 2.05$, $p < .05$; all other conparisons were insignificant. Next, we analyzed the high incentive data.

Neither the Standards effect at high incentive, $F(3, 262) < 1.00$ $p > .05$, nor the Expressiveness x Standards interaction at high incentive, $F(3, 262) < 1.00$ $p > .05$, was significant.

There are two points to be made about Experiment 2. First, the low incentive findings partially replicated the findings of Experiment 1. For high expressive instructors both experiments showed that giving higher grades could sometimes improve ratings. However, for low expressive instructors, the results of the two experiments were not compatible. Second, we had expected grading standards to have a large impact on ratings with high student incentive. Surprisingly, there was no effect of grading standards at high incentive.

General Discussion

The results of our two experiments suggest that some instructors may receive different ratings based only on the grading standards they employ. But the relationship between grading standards and ratings is not clearcut. Our studies suggest that under certain conditions, instructors who assign high grades can improve their ratings while under other conditions high grades will either adversely affect ratings or will not affect them at all. Because of the many differences between Experiment 1 and the low incentive conditions of Experiment 2, we are not sure why the results of the two studies were inconsistent. Differences in lecture topics, identity of the lecturer, degree of instructor expressiveness, videotape presentation format, etc. might account for the discrepancies. These differences highlight the complexity of the phenomenon. We were especially surprised to find no effect of grading standards at high incentive. We had expected that for high incentive students, the

effect of grades would be large. To us, the high level of student incentive made our laboratory situation more like typical college classes. Thus, our data suggest that for typical college classes, where students have strong incentives to learn, grading standards do not affect ratings. When student incentive is not as strong, grading standards may affect ratings.

To the extent that our high incentive conditions are representative of typical classes, we failed to find data to support Snyder and Clair's (1976) claim that TRFs are not valid because grades affect ratings. Under low student incentive, we found that grading standards may sometimes affect ratings and, therefore, found support for some TRF critics. However, unlike prior claims, easy graders do not always receive higher ratings than strict graders; sometimes easy graders may receive lower ratings.

References

Dixon, W. J. (Ed.). BMDP biomedical computer programs. Berkeley: University of California Press, 1975.

Feldman, K. A. Grades and college student's evaluations of their courses and teachers. Research in Higher Education, 1976, 4, 69-111.

Hays, W. L. Statistics for the social sciences (2nd Ed.). New York: Holt, Rinehart, and Winston, 1973.

Kirk, R. Experimental design: Procedures for the behavioral sciences. Belmont, Cal.: Brooks/Cole, 1968.

Leventhal, L., Abrami, P. C., and Perry, R. P. Bogus evidence for the validity of student ratings. Paper presented at the annual meeting of the American Psychological Association, San Francisco, August 1977. (ERIC Document Reproduction Service No. ED 150 510)

Pohlmann, J. T. A multivariate analysis of selected class characteristics and student ratings of instruction. Multivariate Behavioral Research, 1975, 10, 81-92.

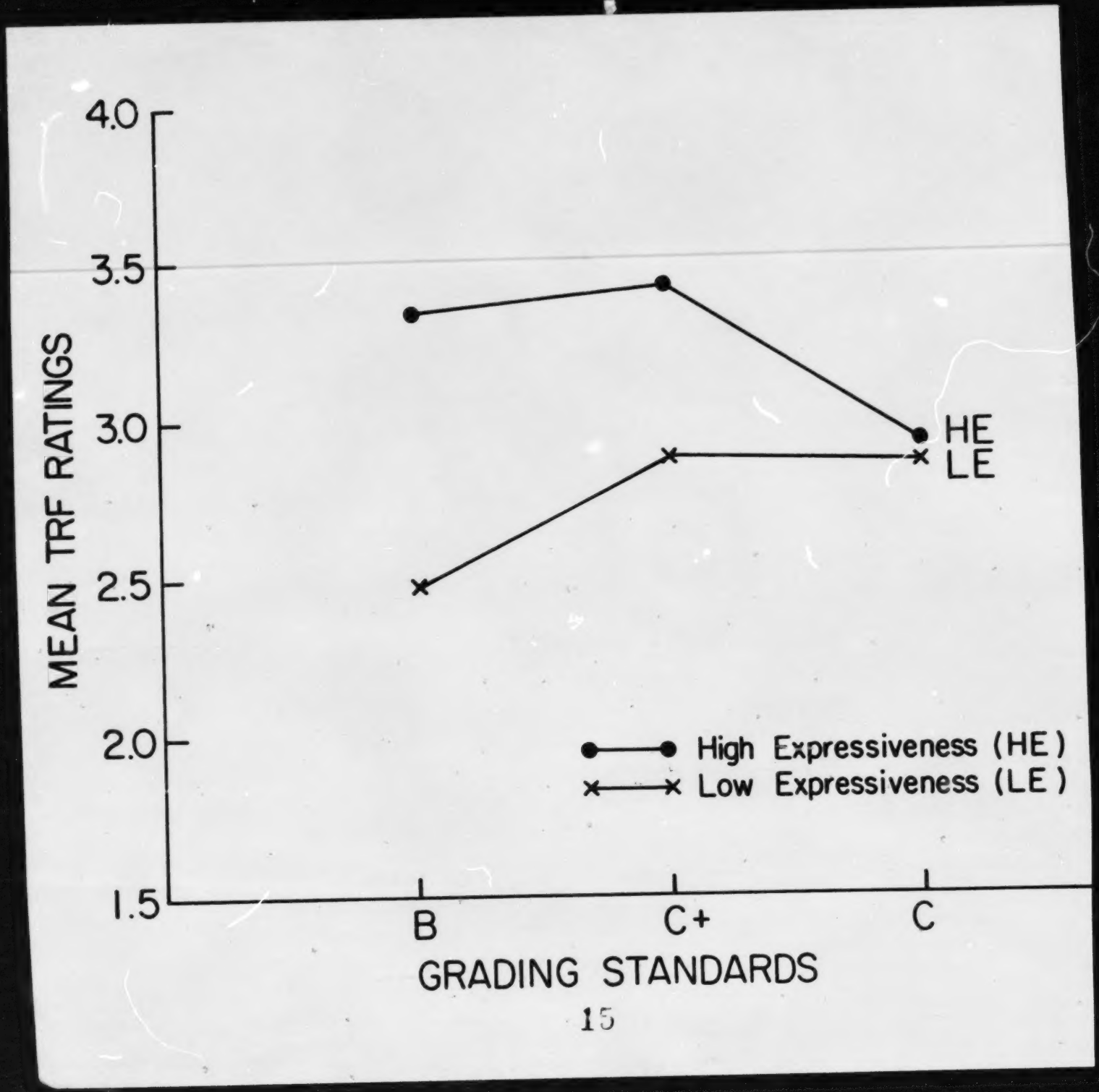
Snyder, C. R., & Clair, M. Effects of expected and obtained grades on teacher evaluation and attribution of performance. Journal of Educational Psychology, 1976, 68, 75-82.

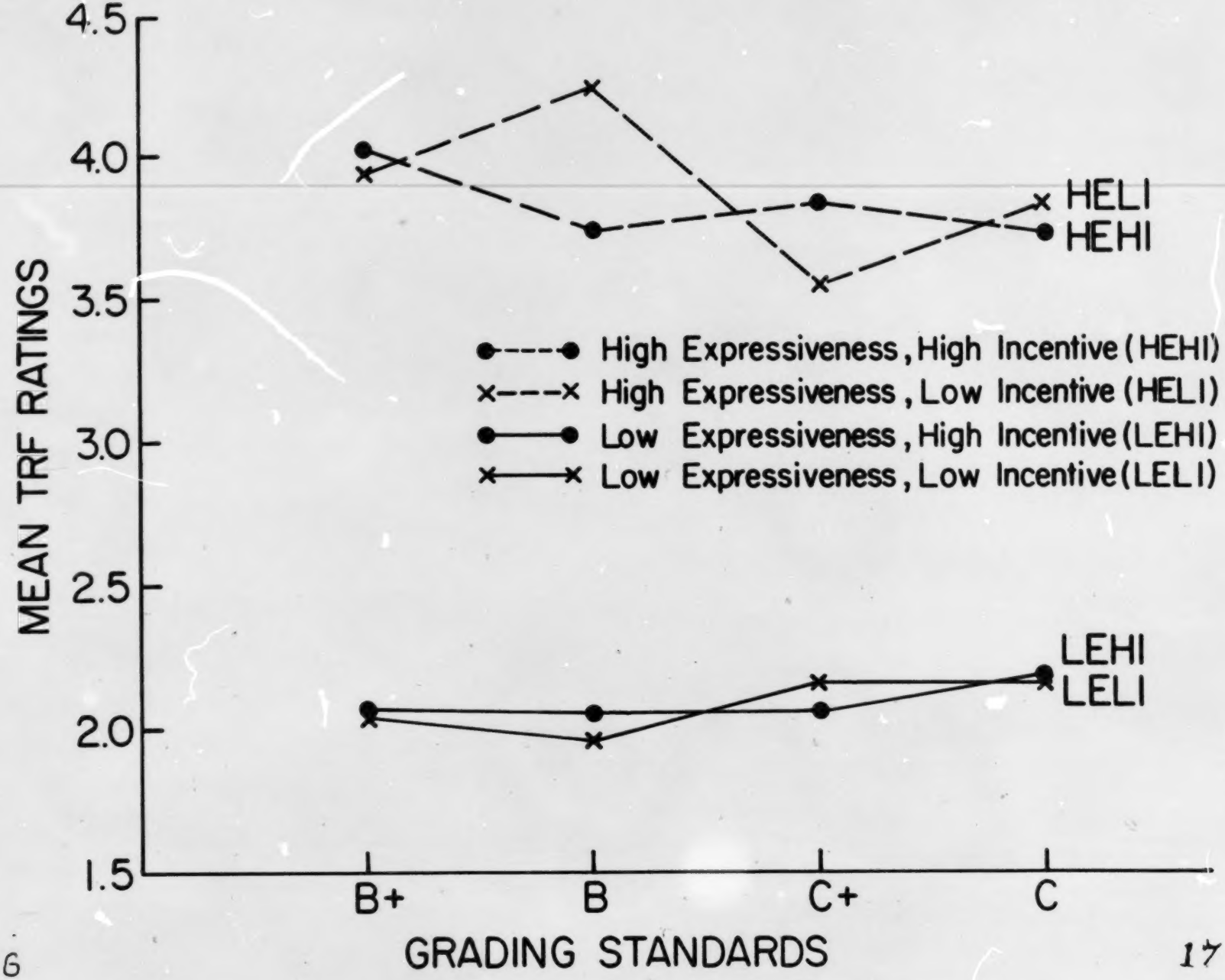
Williams, R. G., & Ware, J. E. An extended visit with Dr. Fox: Validity of student satisfaction with instruction ratings after repeated exposures to a lecturer. American Educational Research Journal, 1977, 14, 449-458.

Figure Captions

Figure 1. Mean TRF in Experiment 1 for low expressiveness and high expressiveness students under different grading standard conditions.

Figure 2. Mean TRF ratings in Experiment 2 for low expressiveness and high expressiveness students under different grading standard and incentive conditions.


4.0
3.5
3.0
2.5
2.0
1.5
MEAN TRF RATINGS
HE
LE
High Expressiveness (HE)
Low Expressiveness (LE)
B
C+
C
GRADING STANDARDS


4.5
4.0
3.5
3.0
2.5
2.0
1.5
MEAN TRF RATINGS
B+
B
C+
C
GRADING STANDARDS
HELI
HEHI
LEHI
LELI
High Expressiveness, High Incentive (HEHI)
High Expressiveness, Low Incentive (HELI)
Low Expressiveness, High Incentive (LEHI)
Low Expressiveness, Low Incentive (LELI)